Error bounds on the power method for determining the largest eigenvalue of a symmetric, positive definite matrix

Joel Friedman

Department of Mathematics, University of British Columbia, Vancouver, BC, Canada, V6T 1Z2

Received 1 October 1995; received in revised form 27 January 1998; accepted 6 February 1998

Submitted by R.A. Brualdi

Abstract

Let $A$ be a positive definite, symmetric matrix. We wish to determine the largest eigenvalue, $\lambda_1$. We consider the power method, i.e. that of choosing a vector $v_0$ and setting $v_k = A^k v_0$, then the Rayleigh quotients $R_k = (Av_k, v_k)/(v_k, v_k)$ usually converge to $\lambda_1$ as $k \to \infty$ (here $(u, v)$ denotes their inner product). In this paper we give two methods for determining how close $R_k$ is to $\lambda_1$. They are both based on a bound on $\lambda_1 - R_k$ involving the difference of two consecutive Rayleigh quotients and a quantity $\omega_k$. While we do not know how to directly calculate $\omega_k$, we can give an algorithm for giving a good upper bound on it, at least with high probability. This leads to an upper bound for $\lambda_1 - R_k$ which is proportional to $(\lambda_2/\lambda_1)^k$, which holds with a prescribed probability (the prescribed probability being an arbitrary $\delta > 0$, with the upper bound depending on $\delta$). © 1998 Elsevier Science Inc. All rights reserved.

1. Introduction

Let $A$ be an $n \times n$ symmetric, positive definite matrix, with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. We wish to determine $\lambda_1$. 

1 E-mail: jf@math.ubc.ca.
In this paper we study the usual power method. Namely, we start with a
vector \( v_0 \), and iteratively calculate \( v_k = A^k v \). Then we calculate the Rayleigh quotient
\[
R_k = (Av_k, v_k)/(v_k, v_k) = (v_{k+1}, v_k)/(v_k, v_k),
\]
where \((\cdot, \cdot)\) denotes the inner product. Let \( u_i \) be an orthonormal set of eigenvectors
for \( A \), with \( u_i \) corresponding to \( \lambda_i \), and let \( v_0 = \sum_i c_i u_i \). It is easy to see that
if \( c_1 \neq 0 \), then the limit of \( R_k \) as \( k \to \infty \) is \( \lambda_1 \).

In practice one stops after a finite number of iterations, and hopes that
\( R_k \) gives a good approximation to \( \lambda_1 \). But how good an approximation is it? Of
course, \( R_k \leq \lambda_1 \) gives a lower bound on \( \lambda_1 \), so in this paper we will try to bound
\( \lambda_1 \) from above. Since \( c_1 \) could (theoretically, at least) vanish or almost vanish,
we cannot give a certain upper bound on \( \lambda_1 \). However, we will bound \( \lambda_1 \) from
above in terms of \( R_k \) and a quantity, \( \omega \), defined below. While \( \omega \) is derived from
\( c_1 \) and is therefore unknown, we can bound \( \omega \) from above "with high proba-
bility". We explain this in more detail in what follows.

In practice one might just compute \( e_k = R_k - R_{k-1} \), and when this
is small then conclude that \( R_k \) is near \( \lambda_1 \). To analyze this, consider what happens for
large \( k \). It is easy to see that \( D_k = \lambda_1 - R_k \) is roughly \( C(\lambda_2/\lambda_1)^{2k} \), assuming
\( \lambda_2 < \lambda_1 \) and \( c_1, c_2 \neq 0 \) (see Ref. [1] or Section 3). \(^2\) Hence \( D_k \approx \epsilon_k(\lambda_2/\lambda_1)^2/\beta \),
where \( \beta = 1 - (\lambda_2^2/\lambda_1^2) \). So \( 1/\beta \) is large if \( \lambda_2 \) is near \( \lambda_1 \), and one must really have
a good estimate of \( \beta \) before one can make claims about \( D_k \) being small. Also,
just because \( e_k \) seems to be zero when the calculations are done in, say, double
precision, it does not mean that \( R_k \) will not get closer to \( \lambda_1 \) as \( k \) increases; in-
deed, while \( D_k \approx C(\lambda_2/\lambda_1)^{2k} \), \( \tilde{v}_k = v_k/\|v_k\| \) converges to an eigenvector \( v \) with
\( v - v_k \approx w(\lambda_2/\lambda_1)^k \) for some eigenvector, \( w \), of \( \lambda_2 \). So while the \( R_k \) appears to
have converged, the \( \tilde{v}_k \) may still be changing and future \( R_k \)'s may come much
closer to \( \lambda_1 \). So what seem like reasonable approaches to estimating
\( D_k = \lambda_1 - R_k \) may not be very reliable.

Giving precise upper bounds on \( D_k \) involving the power method requires an-
alyzing the quantity \( \omega = (\sum c_i^2)/c_1^2 = \|v_0\|^2/c_1^2 \). For example, if \( \omega = \infty \), i.e.
\( c_1 = 0 \), then \( R_k \) converges to \( \lambda_2 \), assuming \( c_2 \neq 0 \) (and exact calculations);
similarly, if \( \omega \) is very large, then convergence to \( \lambda_1 \) will be slow (depending on \( \lambda_2 \)
and \( c_2 \)). Since \( v_0 \) is, for the moment, unrestricted, we have no control over \( \omega \).
Yet, we know that a choice of \( v_0 \) with i.i.d. normally distributed (mean 0) co-
efficients leads to an \( \omega \) of size roughly \( n \); in fact, with probability \( \geq 1 - \delta \) we have
\( \omega - 1 \leq n\delta^{-2}2/\pi \) (see Section 4). Thus we have probabilistic control on \( \omega \).

The point of this paper is to give a method that given a \( \delta > 0 \) produces a
bound on \( \lambda_1 \) which holds with probability \( \geq 1 - \delta \); this probabilistic aspect

\(^2\) In general, if \( \lambda_1 \) has multiplicity \( t \) in \( A \), i.e. \( \lambda_1 = \lambda_t \neq \lambda_{t+1} \), then \( D_k \) is roughly \( C(\lambda_{t+1}/\lambda_1)^{2k} \) for
large \( k \), assuming \( c_{t+1} \neq 0 \) and one of \( c_1, \ldots, c_t \) is nonzero.
arises primarily from $\omega$. Such bounds were first given by (see Refs. [2,3]); the bound in Ref. [3] is based on one Rayleigh quotient and $\omega$; the bound in Ref. [2] is based on $s_k = (\|v_k\|/\|v_0\|)^{1/k}$ and $\omega$. The relative error made by these bounds is proportional to $1/k$ for large $k$. In this paper we give a new bound; it is slightly more elaborate (involving $\epsilon_k$, a difference of Rayleigh quotients) than those in Refs. [2,3], and gives much better bounds asymptotically in the number of iterations. Our bound makes a relative error in $\lambda_1$ which is exponentially decreasing in $k$ for large $k$.

Recall that there are a number of ways of giving estimates on $\lambda_1$ which are often used but are not certain nor can be guaranteed to hold with a certain probability. For example, we can use the asymptotics of $D_k \approx C(\lambda_2/\lambda_1)^{2k}$ to try to approximate $C$, $\lambda_1$, and $\lambda_2/\lambda_1$ based on knowing the $R_k$'s. However, it is hard to know how accurate resulting conclusions will be. As another example, there are techniques which give an interval guaranteed to contain some $\lambda_i$ and "probably" $\lambda_1$ (although this probability has yet to be fully analyzed) (see Refs. [4,5]); however, one is not sure that this interval contains $\lambda_1$, and one might wonder about the extent to which one can be sure.

Recall that a number of papers (see Refs. [2,3,6]) give bounds on $\lambda_1$ in terms of $s_k$ or $R_k$, each time a relative error of $\nu$ requiring roughly $k$ iterations with $2k\nu \geq \log(\omega - 1)$; so they yield a relative error of $\nu$ or less with probability $\geq 1 - \delta$ if $k \geq \nu^{-1} \log(n\delta^{-2}/2\pi)/2$. Indeed, in Ref. [2] one is essentially observing that $\lambda_1 \leq \nu^{1/(2k)} s_k$, which is $\leq \epsilon^k s_k$ (i.e. roughly $\leq (1 + \nu) s_k$) if $2k\nu \geq \log \omega$. In Ref. [3] it is shown that $\lambda_1 \leq (1 + \nu) R_k$ provided that a fairly complicated inequality holds, which to first order is $2k\nu \geq \log(\omega - 1) - \log \log(\omega - 1)$. In both cases the bounds on $k$ are shown to be (or are clearly) tight. While these bounds suffice for large $\nu$, i.e. rough estimates of $\lambda_1$, their behavior as $\nu \to 0$ is not very satisfactory. Indeed, they can only guarantee a relative error of roughly $\log(n\delta^{-2}/2\pi)/(2k) \approx C/k$ after $k$ iterations, while we know that a Rayleigh quotient will have a relative error of roughly $C(\lambda_2/\lambda_1)^{2k}$. Assuming we are willing to perform enough operations so that $(\lambda_2/\lambda_1)^{2k}$ is sufficiently small, we should find methods giving relative errors near this asymptotic bound.

In this paper we give an upper bound for $D_k$ based on the difference between two successive Rayleigh quotients, $\epsilon_k = R_k - R_{k-1}$. It is in using this difference that we obtain an essential improvement over the estimates of Refs. [2,3,6]. Our estimate has a number of forms. The weakest involves only $\omega$ and otherwise known quantities; it gives an estimate for $D_k$ of roughly $C(\lambda_2/\lambda_1)^k$, meaning that one needs roughly $2k$ iterations for an estimate equal to the true error after $k$ iterations. So while this is not optimal, it does give an exponentially decreasing estimate in $k$ (as opposed to previous estimates which are inversely proportional to $k$). A much stronger estimate of $D_k$ can be obtained, but it involves $\lambda_2$, which we have not been estimating. So we need some estimate of $\lambda_2$ to apply this estimate; however, having already calculated $v_k$ for a large $k$ enables one to get a fairly good upper bound on $\lambda_2$, assuming $k$ is large enough so that...
\begin{align*}
\lambda_2 &\leq R_k k/(k + 1). \text{ This last condition can be verified to probability } \geq 1 - \delta \text{ in roughly } (2v)^{-1} \log(n\delta^{-2}/\pi) \text{ iterations of a certain power method, with } v = (\lambda_1/\lambda_2) - 1 + O(1/k). \text{ Using this stronger estimate for } D_k \text{ we get a bound that is roughly } \bar{C}(\lambda_2/\lambda_1)^{2k} \text{, i.e. within a constant factor of optimal (see Theorem 3.2). The only price here is the extra iterations of a certain power method needed. However, the constant } \bar{C} \text{ is off from optimal by several factors including a } \delta^{-2}\sqrt{\pi} \text{ factor, and this will hopefully be improved upon in the future.}

1.1. Some bounds on } D_k

Let
\begin{align*}
\sigma_r &= \sum_{i=1}^{n} c_i^2 \lambda_i^r, \\
\omega_r &= \sigma_r/(c_1^2 \lambda_1^r).
\end{align*}

Then
\begin{align*}
\sigma_{2k} &= (v_k, v_k), \\
\sigma_{2k+1} &= (v_{k+1}, v_k),
\end{align*}

and so the } \sigma_r \text{ are computable. Of course, } R_k = \sigma_{2k+1}/\sigma_{2k}. \text{ The } \omega_r \text{ are not readily computable, but we have } \omega = \omega_0 \text{ which we can bound probabilistically. Also, } \omega_{2k} \text{ is to } v_k \text{ what } \omega \text{ is to } v_0, \text{ in that } \omega_{2k} = \|v_k\|^2/c_{1,k}^2, \text{ where } c_{1,k} = c_1 \lambda_1^k \text{ is the } u_1 \text{ component of } v_k.

\textbf{Theorem 1.1. For any } k \geq 2, \text{ we have } \epsilon_k = R_k - R_{k-1} \text{ is necessarily positive, and}
\begin{equation}
D_k \leq \frac{1}{2} \left( f_k + \sqrt{f_k(4R_k + f_k)} \right), \tag{1.1}
\end{equation}

\text{where } f_k = \epsilon_k \omega_{2k-2}(\omega_{2k+2} - 1)/\omega_{2k}.

Both this theorem and the next proposition will be proven in Section 2.

\textbf{Proposition 1.2. We have}
\begin{align*}
\omega_{2k-2}(\omega_{2k+2} - 1)/\omega_{2k} &\leq \omega_{2k-2} - 1 \leq (\omega - 1)(\lambda_2/\lambda_1)^{2k-2} \leq \omega - 1,
\end{align*}

\text{and Theorem 1.1 remains true with } f_k \text{ taken to be } \epsilon_k \text{ times any of the above expressions in between the } \leq \text{'s.}

In Section 3 and Theorem 3.2 we show how this can be used to give an upper bound on } D_k \text{ which is asymptotically roughly } C_6(\lambda_1/\lambda_2)^{2k-2} \text{ for a constant } C_6.

While the above difference of Rayleigh quotients, } \epsilon_k \text{, is a fairly natural quantity to use in bounding the } D_k \text{ from above, the resulting upper bound always loses a factor of roughly } \sqrt{1 + (\lambda_2/\lambda_1)} \text{ for large } k; \text{ for } \lambda_2 \text{ near } \lambda_1 \text{ this is a loss of a factor of roughly } \sqrt{2}. \text{ There are variants of this theorem which modify the Rayleigh quotient difference but do not involve such a loss. We describe them next.
1.2. Variants of the $D_k$ bound

We describe some variants of the main theorem; we will see in Section 3 that they are fairly tight. They will also be proven in Section 2.

The $T_k = \sigma_k + 1 / \sigma_k$ bound: Let $T_k = \sigma_k + 1 / \sigma_k$. The $T_k$ can be computed from the $v_k$ and $A$ as indicated before (for $k$ even $T_k$ is the usual Rayleigh quotient of $v_k / 2$ with respect to $A$). We have the following theorem.

**Theorem 1.3.** For any $k \geq 2$, we have $\epsilon_k = T_k - T_{k-1}$ is necessarily positive, and

$$D_k \leq \frac{1}{2} \left( f_k + \sqrt{f_k (4T_k + f_k)} \right),$$

where $f_k = \epsilon_k \omega_{k-1} (\omega_{k+1} - 1) / \omega_k$.

* A not necessarily positive definite: Let $A$ be symmetric, but not necessarily positive definite. Let its eigenvalues be $|\lambda_1| \geq \cdots \geq |\lambda_n|$. We can determine $|\lambda_1|$ by applying the previous methods to $A^2$. Or we can compute the Rayleigh quotients for $A^2$:

$$S_k = (A^2 v_k, v_k) / (v_k, v_k) = (v_{k+1}, v_{k+1}) / (v_k, v_k) = \sigma_{2k+2} / \sigma_{2k}$$

with notation as before. Let $\epsilon_k = S_k - S_{k-1}$, and let $D_k = \lambda_1^2 - S_k$.

**Theorem 1.4.** For any $k \geq 2$, we have $\epsilon_k$ is necessarily positive, and

$$D_k \leq \frac{1}{2} \left( f_k + \sqrt{f_k (4S_k + f_k)} \right),$$

where $f_k = \epsilon_k \omega_{2k-2} (\omega_{2k+2} - 1) / \omega_{2k}$.

1.3. Bounding $\omega$

The estimates of $D_k$ all involve $\omega$ to a certain extent. In this paper we will most often use the estimate (proven in Section 4):

$$\Pr(\omega - 1 \geq n\alpha) \leq \sqrt{2/(\pi \alpha)},$$

where $\Pr(\ )$ is the probability. This holds for $v_0$ chosen with i.i.d. normal components as mentioned before. Another way of saying this is that for any $\delta > 0$, $\omega - 1 \leq n\delta^{-2} - 2 / \pi$ with probability $\geq 1 - \delta$.

Here we mention some results proven in Section 4, and give some ways to apply them that are different from that in Section 3; for example, by running the same experiment a number of times (with different $v_0$'s), one can dramatically increase the probability that a bound on $\omega$ holds (in at least one of the runs). After describing these bounds we give a guaranteed bound, which requires the algorithm to be run $n$ times (this is more of a philosophical comment); this was suggested to me by Philip Loewen.

In Section 4 we will see proof of the following theorem.
Theorem 1.5. Let the components of \( v_0 \) be chosen independently, each from the normal distribution with 0 mean and variance 1. Then

\[
\Pr(\omega - 1 \geq na) \leq e^{1/(2a^2n)} \text{erf}\left(\frac{1}{\sqrt{2a}}\right).
\]

Here erf is the standard error function,

\[
\text{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-t^2} dt.
\]

The analysis in the above theorem is sharp to within a multiplicative factor of \( 1 + O(1/n) \) for any fixed \( a \). We will, in Section 4, describe the case of other distributions for the components, but our estimates will not be so sharp.

As a corollary to the theorem we have, for example,

\[
\Pr(\omega - 1 > 100n) < 0.07966 e^{1/(20000n)} < 0.08
\]

for all \( n \) positive integers. So if we run the algorithm once, and set \( \omega = 100n + 1 \) in applying the above theorem, we can be sure that our estimate holds with probability \( \geq 1 - 0.08 = 0.92 \). However, if we ran the algorithm 10 times (with 10 different \( v_0 \)'s), and took the largest upper bound on \( \lambda_1^2 \) given by the theorem with \( \omega = 100n + 1 \), this upper bound would hold with probability \( \geq 1 - (0.08)^{10} \geq 1 - 1.074 \times 10^{-11} \).

If we desired a tighter upper bound on \( \lambda_1 \), we might only be able to tolerate a bound of \( \omega \leq n + 1 \). We then have

\[
\Pr(\omega - 1 \geq n) \leq 0.6827 e^{1/(2n)};
\]

here we would need more runs of the power method to guarantee with high probability that in one of the runs we have \( \omega \leq n + 1 \).

We finish this section by remarking that if one is willing to make \( n \) runs of the power method, then \( \omega \) can be bounded above in at least one of the runs. Simply take \( v_0 \) running through the standard basis \( e_1 = (1,0,\ldots,0), \ldots, e_n = (0,\ldots,0,1) \). Then since \( u_1 \) has a component, say the \( i \)th, at least \( 1/\sqrt{n} \), we get \( |c_i| \geq 1/\sqrt{n} \) and so \( \omega = |v_0|^2/c_i^2 = |e_i|^2/c_i^2 \leq n \).

1.4. The rest of this paper

In Section 2 we prove Theorems 1.1, 1.3, 1.4, and Proposition 1.2. In Section 3 we give an asymptotic analysis of the bounds in Theorems 1.1, 1.3, and 1.4; we concentrate on Theorem 1.1 – the analysis for the others is quite similar. In Section 4 we give some estimates on the probability that \( \omega \) is large, for various distributions. In Appendix A we give the details of a Beta function bound needed in Section 4.
2. Proofs of the main theorems

We begin by explaining Proposition 1.2. We have
\[ \omega_k - \omega_{k+1} = \sum_{i>1} c_i^2 \lambda_i^k (\lambda_i - \lambda_1) \geq 0, \]
and so \( \omega_k \) is decreasing in \( k \). Also
\[ c_1^2 \lambda_k (\omega_k - 1) = \sum_{i>1} c_i^2 \lambda_i^k \leq \left( \sum_{i>1} c_i^2 \right) \lambda_2^k - c_1^2 (\omega - 1) \lambda_2^k, \]
and so
\[ \omega_k - 1 \leq (\omega - 1) (\lambda_2 / \lambda_1)^k. \]
Finally, notice that
\[ \omega_{2k-2}(\omega_{2k+2} - 1) / \omega_{2k} = (\omega_{2k-2} / \omega_{2k}) - (\omega_{2k-2} / \omega_{2k}) \leq \omega_{2k-2} - 1, \]
just using the monotonicity of \( \omega_r \). Hence the proposition follows.

The proofs of Theorems 1.1, 1.3, and 1.4 are quite similar. The latter two are virtually identical, so we start with them. First we prove Theorem 1.4.

**Proof.** We start by observing that
\[
\sigma_{2k+2} \sigma_{2k-2} - \sigma_{2k}^2 = \sum_{i,j} c_i^2 c_j^2 (\lambda_i^{2k+2} \lambda_j^{2k-2} - \lambda_i^{2k} \lambda_j^{2k}) = \sum_{i<j} c_i^2 c_j^2 (\lambda_i^{2k+2} \lambda_j^{2k-2} + \lambda_i^{2k-2} \lambda_j^{2k+2} - 2 \lambda_i^{2k} \lambda_j^{2k}) = \sum_{i<j} c_i^2 c_j^2 \lambda_i^{2k-2} \lambda_j^{2k-2} (\lambda_i^2 - \lambda_j^2)^2.
\]
Therefore
\[
\epsilon_k = (\sigma_{2k+2} / \sigma_{2k}) - (\sigma_{2k} / \sigma_{2k-2}) = \sum_{i<j} \lambda_i^{2k-2} \lambda_j^{2k-2} c_i^2 c_j^2 (\lambda_i^2 - \lambda_j^2)^2 / \sigma_{2k} \sigma_{2k-2} \geq \sum_{i<j} \lambda_i^{2k-2} \lambda_j^{2k-2} c_i^2 c_j^2 (\lambda_i^2 - \lambda_j^2)^2 / \sigma_{2k} \sigma_{2k-2} \sigma_{2k} \sigma_{2k-2} = \sum_{i<j} \lambda_i^{2k-2} c_i^2 c_j^2 (\lambda_i^2 - \lambda_j^2)^2 / \sigma_{2k} \sigma_{2k-2} \sigma_{2k} \sigma_{2k-2}.
\]
(2.1)
Since the above expression in Eq. (2.1) is non-negative, so is \( \epsilon_k \).

Next, notice that
\[
D_k = \lambda_1^2 - (\sigma_{2k+2} / \sigma_{2k}) = \sum_{i} \lambda_i^{2k} c_i^2 (\lambda_i^2 - \lambda_j^2) / \sigma_{2k}
\]
which, via Cauchy–Schwarz, is

\[
\left( \sum_{i>1} \frac{\lambda_i^{2k+2} c_i^2 (\lambda_i^2 - \lambda_j^2)^2}{\sigma_{2k}} \right)^{1/2} \left( \frac{\sum_{i>1} \lambda_i^{2k+2} c_i^2}{\sigma_{2k}} \right)^{1/2}. \tag{2.2}
\]

Clearly we have

\[
\left( \sum_{i>1} \frac{\lambda_i^{2k+2} c_i^2}{\sigma_{2k}} \right)^{1/2} = \left( \lambda_1^2 (\omega_{2k+2} - 1)/\omega_{2k} \right)^{1/2}.
\]

Hence we have

\[
D_k^2 \leq \left( \lambda_1^2 (\omega_{2k+2} - 1)/\omega_{2k} \right) \left( \sum_{i>1} \frac{\lambda_i^{2k-2} c_i^2 (\lambda_i^2 - \lambda_j^2)^2}{\sigma_{2k}} \right)
\leq \left( \lambda_1^2 (\omega_{2k+2} - 1)/\omega_{2k} \right) (\epsilon_k \omega_{2k-2}).
\]

Using \( \lambda_1^2 = D_k + S_k \) (by definition), we have

\[
D_k^2 \leq (D_k + S_k) f_k,
\]

where \( f_k = \epsilon_k \omega_{2k-2} (\omega_{2k+2} - 1)/\omega_{2k} \), which implies

\[
D_k \leq \frac{f_k + \sqrt{f_k^2 + 4S_k f_k}}{2}. \quad \Box
\]

To prove Theorem 1.3 we have that

\[
\epsilon_k = (\sigma_{k+1} \sigma_{k-1} - \sigma_k^2)/(\sigma_k \sigma_{k-1}).
\]

So the entire proof goes over as before, just by replacing all exponents and most subscripts by 1/2 their value in the previous proof.

Finally we give the proof of Theorem 1.1. It is similar, but not identical, to the previous proofs. Here we have

\[
\epsilon_k = (\sigma_{2k+1} \sigma_{2k-2} - \sigma_{2k+1} \sigma_{2k})/(\sigma_{2k} \sigma_{2k-2}),
\]

and will use

\[
\sigma_{2k+1} \sigma_{2k-2} - \sigma_{2k+1} \sigma_{2k} = \sum_{i<j} c_i^2 c_j^2 \lambda_i^{2k-2} \lambda_j^{2k-2} (\lambda_i + \lambda_j)(\lambda_i - \lambda_j)^2.
\]

We restrict the above sum of \( i < j \) to \( i = 1 \), as before; however, now we also use \( \lambda_1 + \lambda_j \geq \lambda_1 \) in the above to conclude

\[
\sigma_{2k+1} \sigma_{2k-2} - \sigma_{2k+1} \sigma_{2k} \geq \lambda_1^{2k-1} c_1^2 \sum_{1<i<j} c_j^2 \lambda_j^{2k-2} (\lambda_1 - \lambda_j)^2.
\]

Notice that the dominant summand is the \( j = 2 \) summand, and if \( \lambda_2 \) is near \( \lambda_1 \) then the \( \lambda_1 + \lambda_j \geq \lambda_1 \) estimate loses roughly a factor of 2. We conclude
We apply Cauchy–Schwarz to match the above expression. That is, we write

\[ D_k = \lambda_1 - \frac{\sigma_{2k+1}/\sigma_{2k}}{\sum_i c_i^2 \lambda_i^{2k-2} (\lambda_1 - \lambda_i)^2}. \]

As before we conclude that \( \delta < 2 \).

3. An asymptotic analysis

In this section we will discuss the estimate of Theorem 1.1; a similar discussion holds for that of Theorems 1.3 and 1.4.

Throughout this section we will assume \( \lambda_1 > \lambda_2 > \lambda_3 \) and that \( c_1 \) and \( c_2 \) are nonzero. If this is not the case, i.e. the first and second largest eigenvalues have multiplicity >1, or some of \( c_1, c_2 \) vanish, one can modify the theory and one gets similar results.

First we note that the bound for \( D_k \)

\[ \frac{1}{2} \left( f_k + \sqrt{f_k (4R_k + f_k)} \right) \leq \frac{1}{2} \left( f_k + \sqrt{f_k (4\lambda_1 + f_k)} \right) \leq f_k + \sqrt{f_k \lambda_1} \]

(using \( \sqrt{a+b} \leq \sqrt{a} + \sqrt{b} \) for \( a, b \geq 0 \)). We will often state things in terms of the relative error,

\[ D_k/\lambda_1 \leq (f_k/\lambda_1) + \sqrt{f_k/\lambda_1}. \quad (3.1) \]

To derive asymptotics, notice that

\[ \sigma_r = \sum_i c_i^2 \lambda_i^r = c_1^2 \lambda_1^r [1 + (\lambda_2/\lambda_1)^r c_2^2/c_1^2 + O((\lambda_3/\lambda_1)^r)], \]

and so

\[ R_k = \lambda_1 \left[ 1 + C(\lambda_2/\lambda_1)^{2k}(1 + o(1)) \right], \]

where \( C = \alpha c_2^2/c_1^2 \) with \( \alpha = 1 - (\lambda_2/\lambda_1) \). Hence we have

\[ D_k/\lambda_1 \approx C(\lambda_2/\lambda_1)^{2k}, \quad \epsilon_k/\lambda_1 = (D_k - D_{k-1})/\lambda_1 \approx (C\beta)(\lambda_2/\lambda_1)^{2k-2}, \]

where \( \beta = 1 - (\lambda_2/\lambda_1)^2 \) (as before), and where we write \( a(k) \approx b(k) \) if \( a(k)/b(k) \to 1 \) as \( k \to \infty \).

For any \( \delta > 0 \), we have \( \omega - 1 \leq n\delta^{-2}2/\pi \) with probability \( \geq 1 - \delta \). Using \( f_k = \epsilon_k(\omega - 1) \), Eq. (3.1) will give an estimate on \( D_k \) of
We conclude:

**Theorem 3.1.** For any \( \delta > 0 \), inequality (1.1) with \( f_k = \epsilon_k(\omega - 1) \) yields an estimate of the relative error, \( D_k / \lambda_1 \), true with probability \( \geq 1 - \delta \), which is \( \approx C'(\lambda_2 / \lambda_1)^k \), where \( C' = \delta^{-1} \sqrt{2C \beta n / \pi(\lambda_1 / \lambda_2)} \) with \( \beta, C \) as before.

So the estimate decreases exponentially in \( k \). However, the true relative error is \( \approx C(\lambda_2 / \lambda_1)^{2k} \), and so we need to take roughly \( 2^k \) iterations to get an estimated error equal to the true error after \( k \) iterations. Put another way, if the estimated error is \( \gamma \), the true error will be \( \approx C'' \gamma^2 \).

Next consider the inequality (1.1) with \( f_k = \epsilon_k \omega_{2(k-2)}(\omega_{2k+2} - 1) / \omega_{2k} \). We have

\[
\omega_{2(k-2)}(\omega_{2k+2} - 1) / \omega_{2k} \approx \left( \epsilon_k^2 / \epsilon_1^2 \right) (\lambda_2 / \lambda_1)^{2k+2},
\]

and so \( f_k / \lambda_1 \approx C_3(\lambda_2 / \lambda_1)^{2k} \), where \( C_3 = C / \alpha \). It follows that this estimate is within a constant factor of optimal, but is off by a factor of \( \sqrt{\beta / \alpha} = \sqrt{(\lambda_1 + \lambda_j) / \lambda_1} \), which can be roughly \( \sqrt{2} \). This is precisely the near factor of 2 discarded in the estimate \( \lambda_1 + \lambda_j \geq \lambda_1 \) used in the proof of Theorem 1.1. Of course, our main problem with applying this estimate is that we do not know of any way of computing (even a good estimate of) \( \omega_{2(k-2)}(\omega_{2k+2} - 1) / \omega_{2k} \) (based on computable quantities).

As a compromise, we consider inequality (1.1) with \( f_k = \epsilon_k(\omega - 1) (\lambda_2 / \lambda_1)^{2k-2} \). To apply this, we require an estimate of \( \lambda_2 \). To do this, first assume that we have verified by some method that \( \lambda_2 \leq R_k k / (k + 1) \). Notice (by differentiation) that the function \( f(x) = x^2(R_k - x)^2 \) is strictly increasing in \( x \) from 0 to \( x = R_k k / (k + 1) \), if \( f(x_0) = B \), then \( x_0 = g_B(x_0) \), where \( g_B(x) = B^{-1/2}(R_k - x)^{-1/k} \). The monotonicity of \( f \) implies that \( x_0, g_B(x_0), g_B(g_B(x_0)), \ldots \) is monotone decreasing with limit \( x_0 \). In particular we have \( x_0 < g_B(g_B(x_0)) \).

We can compute \( v_k = R_k v_k - v_{k+1} \) and

\[
(\bar{v}_k, \bar{v}_k) = \sum_i c_i^2 \lambda_i^{2k} (R_k - \lambda_i)^2.
\]

Clearly we have

\[
c_2^2 \lambda_2^{2k} (R_k - \lambda_2)^2 \leq (\bar{v}_k, \bar{v}_k),
\]

i.e.,

\[
f(\lambda_2) \leq (\bar{v}_k, \bar{v}_k) / c_2^2.
\]
Of course, we do not know what $c_2$ is, but we can bound $1/c_2^2$ probabilistically, concluding from Eq. (3.2) that

$$f(\lambda_2) \leq (\bar{v}_k, \bar{v}_k)\delta^{-2}2/\pi,$$

with probability $\geq 1 - \delta$, in which case

$$\lambda_2 \leq \lambda_2^* = g_\theta(g_\theta(x_\star)),
$$

(3.3)

where $B = (\bar{v}_k, \bar{v}_k)\delta^{-2}2/\pi$.

Let us see what kind of bound on $\lambda_2$ Eq. (3.3) produces. We have

$$B = \lambda_2 = R_k - A^2 = c(R_k - A^2)^2/\pi,$$

where $B = (\bar{v}_k, \bar{v}_k)/c_2^2$ is a good approximation to $f(\lambda_2)$. Since

$$B \approx c_2^2\lambda_2^2(R_k - \lambda_2)^2\delta^{-2}2/\pi,$$

we easily compute

$$g_\theta(x_\star) = \lambda_2 \left(1 + \frac{\log C_4(1 + o(1))\log k}{k}\right),$$

where $C_4 = c_2(R_k - \lambda_2)^2/\pi$, and that

$$g_\theta(g_\theta(x_\star)) = \lambda_2 \left(1 + \frac{\log C_5 + o(1)}{k}\right),$$

where $C_5 = c_2\delta^{-1}\sqrt{2/\pi}$. So we can bound

$$(\lambda_2/\lambda_1)^{k-1} \leq (\lambda_\star/R_k)^{k-1} \approx (\lambda_2/\lambda_1)^{k-1}c_2\delta^{-1}\sqrt{2/\pi}.$$ 

Finally we can estimate $D_k/\lambda_1$ by inequality (1.1) with $f_k = \epsilon_k(\omega - 1)$ for $\omega = R_k/2\delta^{-2}2/\pi$ for $\omega - 1$ (with probability $\geq 1 - \delta$) getting a bound which is

$$\approx \sqrt{f_k/\lambda_1} \approx C_6(\lambda_2/\lambda_1)^{2k-2},$$

where $C_6 = c_2(\delta^{-2}2/\pi)\sqrt{nC/\beta}$.

It is interesting to study $C_6$. The factors where the most is lost is the $\sqrt{n}$ factor and the $\delta^2$ factor. The former factor is introduced in the estimate

---

3 Since $c_2$ is normally distributed with variance 1, its density function is everywhere $\leq (2\pi)^{-1/2}$; hence $|c_2| \leq \alpha$ with probability $\leq 2\alpha(2\pi)^{-1/2}$; setting $\alpha = \delta\sqrt{\pi}/2$ yields what is needed just below.
Although $\omega - 1$ is of size roughly $n$, the correct asymptotic factor in place of $\omega - 1$ is $C_2^2/c_1^2$, which tends to be (more or less, probabilistically speaking) a constant. The $\delta^{-2}$ factor arises because we apply the probabilistic statement twice; an alternative would be to take a larger $\delta$ and to run the experiment a few times. However, the number of extra iterations needed to compensate for the $\sqrt{n}\delta^{-2}$ factor in $C_6$ is, of course, only $\log(1/\delta^2)\log(n\delta^{-2})$ which is roughly $x^{-1}\log(\sqrt{n}\delta^{-2})$.

The final question is how to guarantee that $\lambda_2 \leq \mathcal{R}_k/(k + 1)$. For this find $v_k$, and set $A = PAP$ where $P$ is the projection onto the orthogonal complement of $v_k$. From $R_k = \lambda_1 - D_k$, it is easy to see (and a standard fact) that $A$’s largest eigenvalue is between $\lambda_2$ and $\lambda_2 + D_k$. We can now apply the traditional bound such as in Refs. [2,3] to conclude that after $t$ iterations of the power method on $A$ we will obtain a bound on $\lambda_2$, which holds with probability $\geq 1 - \delta$, which is $\leq (1 + v)(\lambda_2 + D_k)$, provided that $2tv \geq \log(n\delta^{-2}/\pi)$. We wish this bound to be $\leq \mathcal{R}_k/(k + 1)$; for large $k$ this will hold, with $v = (\lambda_1/\lambda_2) - 1 + O(1/k)$.

We summarize our findings in the following.

**Theorem 3.2.** Assume that $k$ is large enough so that $\lambda_2 \leq \mathcal{R}_k/(k + 1)$. Then we may bound $D_k/\lambda_1$ by inequality (1.1) with $f_k = \epsilon_k(n\delta^{-2}/\pi)(\lambda_1/\mathcal{R}_k)^{2k-2}$, obtaining a bound which is $\approx C_6(\lambda_2/\lambda_1)^{2k-2}$, where $C_6 = c_2(\delta^{-2}/\pi)^{1/2}nC/\beta$. This bound will hold with probability $\geq 1 - 2\delta$. We can verify, with probability $\geq 1 - \delta$, that $\lambda_2 \leq \mathcal{R}_k/(k + 1)$ holds, assuming that $(1 + v)(\lambda_2 + D_k) \leq \mathcal{R}_k/(k + 1)$, with $(2v)^{-1}\log(n\delta^{-2}/\pi)$ iterations of the power method applied to $A$ (as described above); we have $v = (\lambda_1/\lambda_2) - 1 + O(1/k)$ for large $k$. The total probability for the $D_k/\lambda_1$ bound holding and the condition on $\lambda_2$ being satisfied is therefore $\geq 1 - 3\delta$.

Let us make a few remarks on the theorem. Since $\mathcal{R}_k/(k + 1)$ is increasing in $k$, once we verify that $\lambda_2 \leq \mathcal{R}_k/(k + 1)$ for $k = k_0$, this continues to hold for all $k > k_0$. Also, we do not need to know what $D_k$ or $v$ is in the inequality $(1 + v)(\lambda_2 + D_k) \leq \mathcal{R}_k/(k + 1)$; we simply perform as many iterations as are needed to bound the largest eigenvalue of $A$ by $\mathcal{R}_k/(k + 1)$ (or abort if we feel $k$ is too small and we have iterated for too long). Also there might be some other way or external mechanism for obtaining a reasonable bound on $\lambda_2/\lambda_1$; if so, we could use such a bound in establishing $\lambda_2 \leq \mathcal{R}_k/(k + 1)$ without extra iterations (or even use the bound directly in $f_k = \epsilon_k(\omega - 1)(\lambda_2/\lambda_1)^{2k-2}$ if the bound is very good). Finally, as mentioned in Section 1, to stop the power method at a point when $\epsilon_k$ is small probably requires some knowledge of $\beta = 1 - (\lambda_2/\lambda_1)^2$ to be able to draw rough conclusions about $D_k$; so other algorithms involving the power method may need to establish an estimate on $\lambda_2$, and we should take this into account when comparing other algorithms to this one.
We finish this section by remarking that the same theorems go through almost verbatim for Theorems 1.3 and 1.4. However, we wish to point out that these theorems are asymptotically tight, in a sense, and that there is no loss of a $\sqrt{1 + (\lambda_2/\lambda_1)} \leq \sqrt{2}$ factor asymptotically as there is in Theorem 1.1. Let us perform the analysis for Theorem 1.3. There we have $D_k = \lambda_1 - T_k \approx \lambda_1 C (\lambda_2/\lambda_1)^k$ with $C$ as before. The difference here is that $\epsilon_k = T_k - T_{k-1} \approx C \alpha (\lambda_2/\lambda_1)^{k-1}$ (before we had $\beta$ replacing $\alpha$). So instead of the estimate of $D_k$ being $\approx D_k \sqrt{\beta/\alpha}$, it is now $\approx D_k \sqrt{\alpha/\alpha} = D_k$. The same analysis goes through for Theorem 1.4.

The $\sqrt{2}$ factor saved in the latter two theorems shows that perhaps those $\epsilon_k$'s are easier to work with than that derived from the usual Rayleigh quotients. However, the $\sqrt{2}$ factor saved in the latter theorems is insignificant compared to the factors introduced when trying to practically bound the $f_k$ factor there by some computable numbers.

4. Probabilistic bounds on $\omega$

In this section we discuss $\omega$'s behavior under a few different models. We first remark that $\omega$ is likely to be proportional to $n$ under many different models of a random $v_0$. Indeed, if $v_0 = (\eta_1, \ldots, \eta_n)$ has components $\eta_i$ which are chosen independently from the same distribution, then, by the central limit theorem, $|v_0|^2 = \sum \eta_i^2$ will be close to $n \sigma^2$ where $\sigma^2 = E\{\eta_i^2\}$. Furthermore, $c_1 = (u_1, v_0) = \sum a_i \eta_i$, where $a_1, \ldots, a_n$ are the components of $u_1$; we can usually show that $c_1^2$ stays away from zero with some probability. Hence $\omega = |v_0|^2/c_1^2$ will be no more than proportional to $n$ for large $n$, but will not be likely to be much smaller.

Now we develop some precise bounds on $\omega$. In the case of normal components we will get quite precise bounds. We also discuss the general case, and apply this to uniform components.

4.1. Normal components

Here we give two bounds for the probability that $\omega$ is large, proving Theorem 1.5 as one of the bounds.

Recall, we consider the case where $v_0$'s components, $\eta_1, \ldots, \eta_n$, are independent and each chosen from a normal distribution of mean 0 and variance 1. From the fact that $u_1, \ldots, u_n$ are an orthonormal basis, it follows that $c_1, \ldots, c_n$ are also independent and each chosen from a normal distribution of mean 0 and variance 1. Hence $\omega - 1 = (\sum_{i>1} c_i^2)/c_1^2$ is distributed according to essentially the $F_{n-1,1}$ distribution (actually, to get a true F-distribution one has to normalize $\omega - 1$ by dividing by $n - 1$) (see Ref. [7]). The density function of $\omega - 1$ is
\[ f(x) = \frac{1}{\beta_n (1 + x)^{n/2}} = \frac{1}{\beta_n} \left( \frac{1}{1 + (1/x)} \right)^{n/2} x^{-3/2}, \]

where \( \beta_n \) is a certain value of the Beta function, namely

\[ \beta_n = B((n - 1)/2, 1/2) = \frac{\Gamma((n - 1)/2)\Gamma(1/2)}{\Gamma(n/2)}. \]

While approximations for \( f \) abound in the literature, we seek a fairly precise upper bound. First we note the following lemma.

**Lemma 4.1.** For all \( n \geq 3 \) we have \( 1/\beta_n \leq \sqrt{n/(2\pi)}. \)

The proof is a routine calculation and is given in Appendix A (a proof can also be found in Ref. [2]). This bound is tight to within a multiplicative factor of \( 1 + O(1/n) \), as can be seen in the appendix.

First we apply a crude upper bound, namely

\[ \frac{1}{1 + (1/x)} \leq 1, \]

(not so bad for \( x \) large) to obtain:

\[ f(x) \leq \sqrt{n/(2\pi)} x^{-3/2} \]

and so

\[ \Pr(\omega - 1 \geq \alpha n) \leq \sqrt{n/(2\pi)} 2/\sqrt{\alpha n} = \sqrt{2/\alpha \pi}. \]

So for \( \alpha = 1 \) we have

\[ \Pr(\omega - 1 \geq n) \leq \sqrt{n/(2\pi)} 2/\sqrt{n} = \sqrt{2/\pi} \leq 0.7979. \]

A more precise (for large \( n \)) bound of this type can be obtained by using

\[ 1/(1 + c) \leq 1 - c + c^2 \leq e^{-c+ c^2} \text{ for } c \in (0, 1). \]

This gives

\[ \left( \frac{1}{1 + (1/x)} \right)^{n/2} \leq \exp \left\{ -\frac{n}{2x} + \frac{n}{2x^2} \right\}. \]

Fixing an \( \alpha > 0 \) we have

\[ \int f(x) \, dx \leq e^{1/(2\pi n^2)} \int e^{-n/2x} x^{-3/2} \sqrt{n/(2\pi)} \, dx = \frac{e^{1/(2\pi n^2)}}{\sqrt{2\pi}} \int_x^{\infty} e^{-1/(2u)} u^{-3/2} \, du. \]

It is not hard to see that

\[ \frac{1}{\sqrt{2\pi}} \int e^{-1/(2u)} u^{-3/2} \, du = \text{erf} \left( \frac{1}{\sqrt{2x}} \right), \]
where erf is the standard error function,

\[ \text{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-t^2} \, dt. \]

Hence we have proven Theorem 1.5.

4.2. Other distributions

We remark that for other ways of choosing \( v_0 \) one can still get some bounds, although not as easily and, perhaps, sharply. For example, if we choose \( v_0 \) by picking \( \eta_1, \ldots, \eta_n \) independently from the same distribution, then as remarked before we have \( |v_0|^2 \) will be close to \( n\sigma^2 \). One can use error bounds in the central limit theorem, such as the Berry–Esseen Theorem (see Ref. [8]) to bound above the probability that \( |v_0|^2 \) will not be close to \( n\sigma^2 \). Then one can try to bound the probability that \( c_1 = (u_1, v_0) \) lies in \( (-\beta, \beta) \) for some \( \beta \); this bound would hopefully be independent of \( n \).

We shall give an estimate for \( \eta_i \) drawn from the uniform distribution over \([-1, 1]\). By the central limit theorem \( \eta_1^2 + \cdots + \eta_n^2 \) should be distributed close to \( n/3 \); however, if we are willing to forego a factor of 3, we can use the obvious bound that \( \eta_1^2 + \cdots + \eta_n^2 \lesssim n \).

Next we must argue that \( c_1 = \eta_1 a_1 + \cdots + \eta_n a_n \), where \( a_1, \ldots, a_n \) are the components of \( u_1 \), stays away from 0. Since \( \sum_i \eta_i^2 = 1 \), some \( \eta_i \) must be \( \geq 1/\sqrt{n} \); it follows that \( c_1 \in [-\beta, \beta] \) is \( \leq \sqrt{n}/\beta \). This is not such a good estimate, as it depends on \( n \).

The following moment argument, described to me by Loren Pitt, gives a bound independent of \( n \). First note that

\[ E\{\eta_i^2\} = 1/3, \quad E\{\eta_i^4\} = 1/5, \]

and so

\[ E\{c_1^2\} = \sum_i a_i^2 E\{\eta_i^2\} = 1/3 \]

and

\[ E\{c_1^4\} = \sum_i a_i^4 E\{\eta_i^4\} + \sum_{i,j} 3a_i^2 a_j^2 E\{\eta_i^2\} E\{\eta_j^2\} \lesssim 1/5 + 1/3 = 8/15. \]

Denote the probability measure of \( c_1 \) by \( dP \). Fixing an \( r \) we have

\[ 1/3 = E\{\eta_i^2\} = \int_{|x| \leq r} x^2 \, dP(x) + \int_{|x| > r} x^2 \, dP(x). \]
We bound the first integral by \( r^2 \), and the second by Cauchy–Schwartz:
\[
\int_{|x| > r} x^2 \, dP(x) \leq \sqrt{\int_{|x| > r} dP(x)} \sqrt{\int_{|x| > r} x^4 \, dP(x)} \leq \sqrt{\int_{|x| > r} dP(x) x^4} \sqrt{8/15}.
\]
It follows that
\[
1/3 - r^2 \leq \sqrt{\Pr(|x| > r)(8/15)}.
\]
Assuming \( 1/3 - r^2 \geq 0 \) we may square both sides to obtain
\[
\Pr(|x| > r) \geq (1/3 - r^2)^2 (15/8).
\]
In particular we have for \( r = 1/4 \),
\[
\Pr(|x| > 1/4) \geq 0.1375,
\]
and so we have
\[
\Pr(\omega - 1 \geq 4n) \leq 0.8625.
\]
We can compare this probability to the case where \( \eta_i \) are distributed normally; there we have the better estimate of
\[
\Pr(\omega - 1 \geq 4n) \leq 0.3830 e^{1/(32n)}.
\]
In the normal case we also have
\[
\Pr(\omega - 1 \geq 4n/3) \leq 0.6136 e^{9/(32n)}.
\]
Had we done more work in the uniform case, we could have proved that \( |\eta_0|^2 \) was close to \( n/3 \), and we would have gotten an upper bound on \( \Pr(\omega \geq 1 + 4n/3) \) of roughly 0.8625; again, the normal case has a better estimate. Perhaps a better analysis of \( c_1 \) in the uniform case would get the \( \omega \) estimates close to those in the normal case.

**Acknowledgements**

The author wishes to thank Jim Varah and Beresford Parlett, who gave valuable comments on a draft of this paper. Other valuable discussions and suggestions were provided by Uri Ascher, Philip Loewen, Cindy Greenwood, Loren Pitt, David Boyd, and Alexander Pruss; the author thanks them as well. The author also wishes to thank the NSERC for supporting this research in part. Finally the author wishes to thank a referee of this paper, who pointed out some literature which was new to the author, and whose comments led to a technically much improved version of this paper; this referee is unknown to the author.
Appendix A. A beta function bound

Here we prove Lemma 4.1. Note that $\Gamma(1/2) = \sqrt{\pi}$. Using Stirling's formula,

$$
\Gamma(x + 1) = \sqrt{2\pi x}(x/e)^x e^{\theta/(12x)}
$$

for a $\theta \in [0, 1]$, we easily get

$$
1/\beta_n = 1/B((n - 1)/2, 1/2) 
\leq \frac{1}{\sqrt{\pi}} \frac{\sqrt{\pi(n - 2)} ((n - 2)/2)^{(n - 2)/2} e^{(n - 2)/2}}{\sqrt{\pi(n - 3)} ((n - 3)/2)^{(n - 3)/2} e^{(n - 2)/2}} e^{1/(6n - 12)} 
\leq \frac{1}{\sqrt{\pi}} \sqrt{\frac{n - 2}{n - 3}} \frac{\sqrt{1/2}}{\sqrt{2\pi}} \left( \frac{n - 2}{n - 3} \right)^{(n - 1)/2} \sqrt{n - 2} \frac{1/e}{\sqrt{1/e}} e^{1/(6n - 12)}.
$$

Since

$$
\left( \frac{n - 2}{n - 3} \right)^{n - 3} \leq e,
$$

we further have

$$
1/\beta_n \leq \frac{1}{\sqrt{2\pi}} \frac{n - 2}{\sqrt{n - 3}} e^{1/(6n - 12)} \leq \left( \frac{\sqrt{n}}{\sqrt{2\pi}} \right) \left( \frac{n - 2}{\sqrt{n(n - 3)}} e^{1/(6n - 12)} \right).
$$

It remains to show that the quantity in big parenthesis on the right is bounded by 1. Indeed,

$$
\frac{n - 2}{\sqrt{n^2 - 3n}} = \sqrt{1 - \frac{n - 4}{n^2 - 3n}} \leq \sqrt{1 - \frac{1}{n}} \leq 1 - \frac{1}{2n} \leq e^{-1/(2n)}
$$

hence the quantity in big parenthesis on the right is $\leq \exp[1/(6n - 12)] + [-1/(2n)]$ and since $6n - 12 \geq 2n$ for $n \geq 3$, we are done.

References


